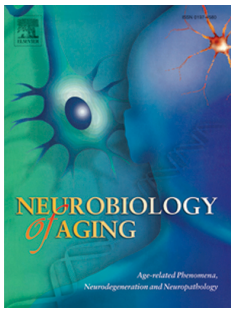
Accepted Manuscript



Heterogeneous patterns of brain atrophy in Alzheimer's disease

Konstantinos Poulakis, Joana B. Pereira, Patricia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen, Simon Lovestone, Andrew Simmons, Lars-Olof Wahlund, Eric Westman

PII: S0197-4580(18)30017-4

DOI: 10.1016/j.neurobiolaging.2018.01.009

Reference: NBA 10135

To appear in: *Neurobiology of Aging*

Received Date: 17 July 2017

Revised Date: 12 January 2018

Accepted Date: 16 January 2018

Please cite this article as: Poulakis, K., Pereira, J.B., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Simmons, A., Wahlund, L.-O., Westman, E., Heterogeneous patterns of brain atrophy in Alzheimer's disease, *Neurobiology of Aging* (2018), doi: 10.1016/j.neurobiolaging.2018.01.009.

Heterogeneous patterns of brain atrophy in Alzheimer's disease

Konstantinos Poulakis[1,*] Joana B. Pereira[1], Patricia Mecocci[2], Bruno Vellas[3], Magda Tsolaki[4], Iwona Kłoszewska[5], Hilkka Soininen[6], Simon Lovestone[7], Andrew Simmons[8,9,10], Lars-Olof Wahlund[1], Eric Westman[1,10,**]

[1] Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden.
[2] Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy.
[3] INSERM U 558, University of Toulouse, Toulouse, France.
[4] 3rd Department of Neurology, Memory and Dementia Unit, Aristotle University of Thessaloniki, Thessaloniki, Greece.
[5] Medical University of Lodz, Lodz, Poland.
[6] University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.
[7] Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK.
[8] NIHR Biomedical Research Centre for Mental Health, London, UK.
[9] NIHR Biomedical Research Unit for Dementia, London, UK.
[10] Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.

* Corresponding authors email: konstantinos.poulakis@ki.se

** Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

There is increasing evidence showing that brain atrophy varies between patients with Alzheimer's disease (AD), suggesting that different anatomical patterns might exist within the same disorder. We investigated AD heterogeneity based on cortical and subcortical atrophy patterns in 299 AD subjects from two multicenter cohorts. Clusters of patients and important discriminative features were determined using random forest pairwise similarity, multidimensional scaling and distance-based hierarchical clustering. We discovered two typical (72.2%) and three atypical (28.8%) subtypes with significantly different demographic, clinical and cognitive characteristics, and different rates of cognitive decline. In contrast to previous studies, our unsupervised random forest approach based on cortical and subcortical volume measures and their linear and non-linear interactions, revealed more typical AD subtypes with important anatomically discriminative features, while the prevalence of atypical cases was lower. The hippocampal-sparing and typical AD subtypes, exhibited worse clinical progression in visuospatial, memory and executive cognitive functions. Our findings suggest there is substantial heterogeneity in AD that has an impact on how patients function and progress over time.

Keywords: Alzheimer's disease; structural magnetic resonance imaging; cortical volumes; subcortical volumes; cluster analyses; Random forest similarity.

1. Introduction

Alzheimer's disease (AD) is a progressive and ultimately fatal neurodegenerative disorder characterized by loss of memory and other cognitive functions (Frisoni et al., 2010). The pathological hallmarks of AD consist of aggregation of β-amyloid (Aβ) peptides into plaques and phosphorylation of tau into neurofibrillary tangles (Shaw et al., 2009), which are thought to spread in the brain in an orderly fashion (Braak and Braak, 1991). For instance, neurofibrillary tangles first appear in the entorhinal cortex and hippocampus, then spread to neocortical association areas and finally reach the primary cortex. However, recent evidence suggests that many patients with AD do not show this spatial pattern of progression of tangles, indicating that there might be distinct pathological subtypes within AD.

In a previous neuropathological study, Murray et al. (2011) found three distinct AD pathological subtypes based on the distribution of tangles in the brain. These subtypes were characterized by either a high number of neurofibrillary tangles in the hippocampus, cortical regions or both and were labeled as limbic-predominant, hippocampal-sparing and typical, respectively. Followed by this description of subtypes, Whitwell et al. (2012) investigated the gray matter differences on structural MRI between these pathological groups and found an association between neurofibrillary tangle deposition and brain atrophy. In addition, the fact that gray matter changes mirrored the regional distribution of tangles suggests that the different pathological subtypes of AD can be investigated *in vivo* using structural neuroimaging.

Up to date, two main classes of approaches have been applied to assess the neuroanatomical heterogeneity in AD. The first class consists of identifying different AD subtypes using a supervised approach based on prior clinical, pathological or

neuroimaging criteria, followed by univariate statistical analysis to define distinct anatomical patterns (Whitwell et al., 2012; Byun et al., 2015; Ferreira et al., 2017). These methods do not allow exploring the variability within the AD population in an unsupervised, unbiased way with respect to its neuroanatomical properties. Moreover, univariate analyses do not consider interactions between variables.

The second class of approaches for the identification of AD subtypes includes unsupervised learning methods such as clustering (Noh et al., 1014; Dong et al., 2016; Hwang et al., 2016; Zhang et al., 2016; Park et al., 2017) and semi-supervised (Varol et al., 2017) multivariate methods using voxel-based or surface-based morphometry measures. These studies found diverse clusters of atrophy that were partially similar to the ones previously reported in the literature (Murray et al., 2011; Whitwell et al., 2012). However, in some studies (Noh et al., 1014; Hwang et al., 2016), the cluster analyses were carried out on a large number of anatomical measures (78,570 cortical thickness vertices) in relatively small groups of AD patients (152 and 77 patients respectively). Carrying out cluster analyses in such high-dimensional datasets of anatomical measures when only small samples of subjects are available can make the distance measures improper to distinguish different groups (Parsons et al., 2004). Moreover, two studies assessed how different subtypes of brain atrophy affected clinical progression in AD (Dong et al., 2016; Na et al., 2016). The first study used both MCI and AD patients in their clustering analyses, which could have influenced the identification of subtypes since not all MCI patients convert to AD (Dong et al., 2016). The second study (Na et al., 2016) assessed the clinical progression of previously defined subtypes (Noh et al., 2014) that we discussed above and it is important to remark that the cluster analysis was based only on cortical thickness measures and did not include subcortical brain structures as in Park et al. (2017).

Hence, to address these limitations, in this study we performed cluster analyses on the gray matter volumes of 162 cortical and subcortical regions of interest (ROI) in 299 AD patients. We applied the Random forest method since it accounts for linear and non-linear interactions between brain regions, which have only been considered in one previous study (Varol et al., 2017). Moreover, in our analyses we did not make any prior assumptions on the number of subtypes or their anatomical patterns in order to avoid inference and interpretation limitations. After defining clusters of brain atrophy, we compared the clinical characteristics between different AD subtypes at baseline and after one year. In addition, we assessed the ability of the hippocampus to cortex volume ratio in discriminating AD groups to evaluate its utility as a marker of heterogeneity in AD, as was suggest by Whitwell et al. (2012). We hypothesized that similar patterns of brain atrophy would be found compared to those described in previous neuropathological studies (Murray et al., 2011; Whitwell et al., 2012). Moreover, we also predicted that these patterns would be associated with distinct cognitive profiles, highlighting the clinical relevance of studying the heterogeneity present in AD.

2. Materials and methods

2.1. Subjects

The dataset used in this study was obtained from two large multicentre cohorts, the AddNeuroMed and the Alzheimer's disease Neuroimaging Initiative (ADNI). AddNeuroMed is an integrated project that is part of InnoMed (the Innovative Medicines Initiative) and funded by the European Union Sixth Framework program (Lovestone et al., 2007, 2009). The main objective of AddNeuroMed is to identify biomarkers or experimental models that can improve diagnosis, prediction and

monitoring of disease progression in AD. Regarding neuroimaging, AddNeuroMed uses magnetic resonance imaging (MRI) and magnetic resonance spectoscopy (MRS) to extract valuable information for the identification of AD biomarkers (Westman et al., 2011). The MRI data for AddNeuroMed was collected from different centres across Europe (Lovestone et al., 2009; Simmons et al., 2009, 2011): University of Perugia (Italy), King's College London (United Kingdom), Aristotle University of Thessaloniki (Greece), University of Kuopio (Finland), University of Lodz (Poland) and University of Toulouse (France). AD patients were recruited from the memory clinics of each of the six participant sites using the following inclusion criteria: 1) fulfilling with the ADRDA/NINCDS for probable AD and DSM-IV criteria for dementia AD; 2) mini-mental state examination (MMSE) score between 12 and 28; 3) age 65 years or above. Exclusion criteria for AD patients consisted of any major unstable systematic organ failure, and significant psychiatric or neurological diseases apart from AD. Controls from AddNeuroMed were recruited from social centers for the elderly or GP surgeries, caregiver's or unrelated family relatives using the following inclusion criteria: 1) MMSE score between 24 and 30; 2) age 65 years or above; 3) geriatric depression scale equal to or less than 5; 4) stable medication; 5) good general health; 6) a clinical dementia rating (CDR) score of zero. Exclusion criteria for controls consisted of fulfilling criteria for dementia (DSM- IV), significant neurological, psychiatric or unstable systematic illness and organ failure.

The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Similarly to the AddNeuroMed study, ADNI strives to reveal sensitive

and specific markers of AD progression in patients from various sites to support the development of new treatments and monitor their effectiveness, as well as to reduce the expenditures of clinical trials. The inclusion criteria for AD patients from ADNI were the following: 1) fulfilling the NINCDS/ADRDA criteria for probable AD; 2) MMSE score between 20 and 26; 3) CDR score of 0.5 or 1.0. Exclusion criteria comprised history of structural brain lesions or head trauma, significant neurological disease other than incipient AD, and the use of psychotropic medications that could affect memory. The inclusion criteria for controls were: 1) MMSE between 24 and 30; 2) CDR score equal to zero. Presence of depression, mild cognitive impairment and dementia were used as exclusion criteria for this group. For more information on the ADNI study, see www.adni-info.org.

Informed consent was obtained from all subjects included in this study. For AD patients, consent was obtained both from the patient and a relative. In the AddNeuroMed cohort, the Alzheimer's disease assessment scale (ADAS) was substituted by the CERAD cognitive battery to assess cognitive functions in controls. The only comparable measure between CERAD (The Consortium to Establish a Registry for Alzheimer’s Disease) and ADAS cognitive batteries is the word recall task (Westman et al., 2011). Therefore the comparison between AD subtypes and controls regarding the ADAS subtests is limited to the word recall subtest.

2.2. MRI acquisition

The MRI data from ADNI and AddNeuroMed were acquired using the same protocol (Jack et al., 2008), which consisted of a high-resolution sagittal 3D T1-weighted MPRAGE volume (voxel size 1.1×1.1×1.2 mm^3) (Westman et al., 2011). Full brain and scull coverage was required for both cohorts, and image quality control took

place according to the AddNeuroMed's quality control procedure (Simmons et al., 2009, 2011). Similarly to previous studies (Pereira et al., 2016; Spulber et al. 2013; Falahati et al. 2016), the data from both cohorts were combined since they show similar patterns of atrophy and predictive power in discriminating AD and MCI patients from controls (Westman et al. 2011).

2.3. MRI preprocessing

T1-weighted images were preprocessed using FreeSurfer (version 5.3; http://freesurfer.net/), which provides cortical and subcortical measures of gray matter volume that can be used for statistical analyses. All data was preprocessed through the HiveDB database system (Muehlboeck et al., 2008). The FreeSurfer preprocessing stream consists of several steps: correction of motion artefacts and spatial distortions; removal of non-brain tissue (Segonne et al., 2004); transformation to the Talairach standard space; intensity normalization (Sled et al., 1998) segmentation of subcortical white matter and deep gray matter structures; tessellation of the gray/white matter boundary; automated topology correction (Segonne et al., 2007); and surface deformation to place the gray/white and gray/CSF borders (Fischl and Dale, 2000). Once the cortical models were complete, registration to a spherical atlas took place, which utilizes individual cortical folding patterns to match cortical geometry across subjects (Fischl et al., 1999). This was followed by parcellation of the cerebral cortex into 148 cortical regions (Destrieux et al., 2010). In this study, we extracted the volumes from these cortical regions in addition to seven gray matter subcortical structures for each hemisphere: hippocampus, amygdala, thalamus, caudate, putamen, accumbens and pallidum (Fischl et al., 2002). We also obtained the estimated total intracranial volume (ICV), an indirect measure of head size, from FreeSurfer

(Buckner et al., 2004). In this study we used volume measures because they are available for cortical and subcortical regions in contrast to cortical thickness, which is not available for subcortical structures in FreeSurfer.

2.4. Cluster analyses

In total, 162 volume measures (148 cortical and 14 subcortical) (Supplementary Table 1 and 2) were included in the cluster analyses to assess different patterns of brain atrophy in AD patients. To correct these volume measures for head size, we applied a residual approach using previously established methods (O'Brien et al., 2006; Voevodskaya et al., 2014). The adjusted volumes were obtained as follows:

$$Volume_{adj} = Volume - b(ICV - \overline{\mathrm{ICV}}),$$

where $Volume_{adj}$ is the ICV-adjusted volume of interest (VOI), $Volume$ is the original uncorrected VOI, b is the slope calculated from the linear regression of $Volume$ of each brain region to the ICV in controls, ICV is the subject's ICV, and $\overline{ICV}$ is the mean ICV across all subjects. We performed cluster analyses on the adjusted 162 regional volumes of AD patients from ADNI and AddNeuroMed using the random forest method (Breiman, 2001) implemented in R (The R Foundation for Statistical Computing; version 3.2.3). The random forest method is an ensemble classifier consisting of many decision trees (Breiman, 1996), where the final result is obtained by combining the predictions of all individual trees. This method combines random feature selection (Amit and Geman, 1997) and bootstrap aggregation (Breiman, 1996), which is important to prevent data overfitting and increase the prediction power.

The random forest architecture allows the computation of a similarity (proximity) measure between pairs of observations (Shi and Horvath, 2006). If we assume

observations i and j, their similarity measure s_{ij} initializes in zero. Each tree then *decides* to assign observations together on a certain class by directing them on the same terminal node. Each time the pair i and j end up on the same terminal node, the similarity measure s_{ij} increases by one. At the end of the process, the similarities are symmetrized and divided by the number of trees. If a dataset has N observations, the resulting pairwise similarity matrix consists by N*N elements. We chose to use the random forest similarity measure instead of a conventional distance measure (e.g. Euclidean) due to the following reasons: a) it carries information regarding the linear and nonlinear interactions of the variables under assessment; b) it provides measures of variable importance; and c) it carries no cluster distribution assumptions.

The random forest classifier distinguishes *synthetic* from observed data if the latter has inherent clusters (Shi and Horvath, 2006). As observed data we used the AD dataset, while as synthetic data we used a dataset that was drawn from a reference distribution such that there is no relationship between observations or variables (Shi and Horvath, 2006). The *Addcl1* method was used in the current study for constructing a synthetic dataset (Shi and Horvath, 2006). Using this method, first we randomly sampled one value from the spectrum of observed values in the first variable of the dataset. The same process was repeated for all the variables resulting in one synthetic observation with in-dependency over variables. A synthetic class was defined by combining the observed and synthetic datasets and labeling them as class 1 and 2, respectively. In order to group patients into clusters, we applied the random forest algorithm to the dataset and extracted the similarity matrix. Additionally, we applied classical multidimensional scaling (MDS) to the similarity matrix to extract a lower dimensional representation of the similarity matrix since the *Addcl1* random forest similarity performs well in conjunction with the MDS (Shi and Horvath, 2006;

Gray et al., 2013). Through that step we were able to visually inspect the pairwise similarity of the subjects in the Euclidean space as well as observe which clusters were sparse. Finally, we applied an agglomerative hierarchical clustering algorithm with average linkage (Gray et al., 2013). This hierarchical clustering method starts by assigning every subject to one cluster. The average linkage uses the mean distance between elements of each cluster as a criterion for merging two clusters together. The algorithm continues merging subjects into clusters until all the subjects form a single group. Then, the desirable cut off (number of clusters) can be decided by using a clustering evaluation criterion or by inspecting the low dimensional coordinates of the data. In this study, we assessed the clustering quality for different number of clusters with six indices and finally used the average Hopkins statistic (Banerjee and Dave, 2004) to decide the cut off value (Supplementary material 2.2). The Hopkins statistic assessed clustering tendency where values between 0 and 0.5 reflect heterogeneous clusters, values around 0.5 reflect clusters that are homogeneous between themselves, and values greater than 0.5 reflect very well defined clusters.

The random forest algorithm induces perturbations in the subject population by design through the bootstrapping feature. Every tree of the forest uses only a subset of the data, which consists of 70% of the original dataset randomly chosen (with replacement) and then the model is validated in the remaining 30%. To further assess the reproducibility of the results, we repeated the random forest 100 times with random seeding and computed the average change in the similarity measures. More information about the random forest parameterization, its evaluation, reproducibility, and the clustering validity can be found in the supplementary material 2.1 and 2.2.

2.5. Supervised analysis

After extracting the clusters, a classification decision tree was built (Breiman, 1996) using the 10 most important anatomical variables from the random forest analysis. We selected 10 variables in order to keep the complexity of the decision tree low and increase its interpretability. Additional variables with lower importance could also be included but this may introduce overfitting and produce discrimination rules with very low confidence for generalization (Müller and Guido, 2017). In this tree, the anatomical variables were included as predictors with no scaling (Müller and Guido, 2017) and the cluster number was included as the dependent variable. In this way, we obtained information regarding how the important anatomical features of the random forest differed for each subtype.

2.6. Statistical analyses

To compare the baseline demographic and clinical characteristics between AD subtypes, we used χ^2 tests for nominal variables and the Kruskal-Wallis test for ordinal variables. To assess differences in cognitive variables between baseline and follow-up within each AD group, the Wilcoxon rank sum test was used. A Holm-Šídák correction was used to control for multiple comparisons in all these analyses.

To calculate the hippocampus to cortex volume ratio, we averaged the left and right hippocampus and divided it by the average of three cortical regions (Whitwell et al., 2012) from the Destrieux atlas: 1) frontal ROI (sum of the bilateral middle frontal cortices), 2) parietal ROI (sum of the bilateral inferior parietal cortices), 3) temporal ROI (sum of the bilateral superior temporal cortices).

3. Results

3.1. AD anatomical subtypes

The optimal number of clusters for the dataset of 162 anatomical variables in 299 AD patients was five, as a function of the average Hopkins statistic (See supplementary material 2.2). Figure 1A shows the random forest similarity matrix after multidimensional scaling and Figure 1B shows the hierarchical clustering tree. At the 5-cluster level, patients were assigned into the following groups with respect to their cortical and subcortical atrophy patterns: (1) a group with minimal brain atrophy in the left entorhinal cortex (minimal atrophy subtype; n= 54; 18.1%), (2) a group with atrophy in temporal and limbic areas (limbic-predominant subtype; n=12; 4%), (3) a group with atrophy mainly in parietal and frontal areas (hippocampal-sparing subtype; n=17; 5.7%), (4) a group with diffuse atrophy in several cortical and subcortical regions, except the postcentral, precentral, caudal middle frontal, paracentral areas and cuneus (diffuse 1 subtype; n=167; 55.8%) and (5) a group with the most severe and widespread brain atrophy in almost all cortical and subcortical regions (diffuse 2 subtype; n=49; 16.4%). Figure 2 shows these different patterns of brain atrophy displayed on the cortical surface in each AD subtype compared to controls, after controlling for age, gender, education and estimated total intracranial volume (eTIV). The differences between groups in subcortical volumes can be found in Table 1. Since two diffuse subtypes were identified in the current study, we compared the patients in these two groups to assess their anatomical differences (Supplementary material 3). This comparison showed increased brain atrophy in frontal, parietal and temporal areas in the diffuse 2 with respect to the diffuse 1 subtype. Using Freesurfer, we carried out a vertex-wise analysis (general linear model) to assess whether the MA subtype showed cortical thinning after one year of follow up, compared to the CN group. No significant results were found.

There was a similar proportion of ADNI and AddNeuroMed patients in each cluster (table 2) and their cortical atrophy patterns were similar (supplementary material 4). The results of the random forest algorithm were similar when applied to ADNI or AddNeuroMed cohorts separately (supplementary material 5).

3.2. Demographic and clinical characteristics in AD subtypes

At baseline, we found that patients with a limbic-predominant or diffuse subtype were significantly older compared to the hippocampal-sparing and minimal atrophy groups ($p < 0.05$). In addition, the limbic-predominant subtype had longer disease duration and the hippocampal-sparing and minimal atrophy subtypes had a younger age at onset of AD compared to the other groups. Education was significantly lower in the group with widespread and diffuse atrophy but only when compared to hippocampal-sparing patients (Table 2).

Regarding cognitive variables, MMSE scores were significantly lower in the diffuse 1 subtype compared to minimal atrophy and diffuse 2 groups and in the diffuse 2 subtype compared to the minimal atrophy group. There were also significant differences in CDR, with the minimal atrophy patients showing lower scores compared to the other groups (Table 2).

At one-year follow-up, there was a significant decline in memory (word recall) and visuospatial (orientation) functions in the hippocampal-sparing subtype. The diffuse 1 group also showed worse cognitive performance in executive (following commands) and visuospatial (ideational praxis, orientation) functions. The diffuse 2 group showed a significant decline in memory (word recognition) and visuospatial functions (orientation, ideational praxis).

3.3. Hippocampus to cortex volume ratio

The limbic-predominant and the minimal atrophy subtypes showed similar values of cortical volume (Figure 3A), but different values of hippocampal volume (Figure 3B). A similar pattern was observed in the diffuse 2 and hippocampal-sparing subtypes, which showed similar cortical volumes (Figure 3A), but differed in hippocampal volumes (Figure 3B).

The values of the hippocampus to cortex volume ratio (Figure 3C) provided a good separation between the limbic-predominant, hippocampal-sparing and diffuse subtypes. However, this ratio was not able to distinguish the minimal-atrophy subtype from the two diffuse subtypes.

Compared to controls, all the AD subtypes had lower hippocampal volume. The cortical volume of the minimal atrophy subtype was similar to controls and the limbic-predominant subtype had similar cortical values to controls. The hippocampus to cortex ratio of the hippocampal-sparing subtype was similar to controls.

3.4. Supervised analysis

The supervised classification tree (Figure 4) showed that the precuneus was the most important region to discriminate AD subtypes. This brain region was on the top of the tree and separated two branches of subtypes: one that included the minimal atrophy, limbic-predominant and diffuse 1 and the other included the hippocampal-sparing and diffuse 2 subtypes. The first branch or group of subtypes had significantly higher volume of the precuneus compared to the second branch of subtypes ($p < 0.0001$, after correction for multiple comparisons). In the first branch, the volumes of the middle occipital gyri separated the minimal atrophy (high values) from the limbic-predominant and diffuse 1 subtypes (low values). In the second branch, the volumes

of the orbital gyri separated the hippocampal-sparing (high values) from the diffuse 2 subtype (low values). Other variables such as the superior temporal, precentral and superior parietal gyri contributed less to the discrimination of the first branch of subtypes. The parietal superior together with the superior occipital gyri were less, but still important variables regarding the discrimination of the diffuse 2 from the hippocampal-sparing subtypes. A complete view of the variable discrimination rules between subtypes is available through the tree representation of the model (Figure 4).

4. Discussion

In this study, we assessed the heterogeneity in the patterns of cortical and subcortical brain atrophy in AD. Our main findings revealed that AD patients could be divided into a minimal atrophy, limbic-predominant, hippocampal-sparing, and two diffuse atrophy subtypes. The five groups differed in age, onset of disease, education and cognitive functions. In addition, the hippocampal-sparing and diffuse subtypes showed worse cognitive performance over time. This suggests that they might be more aggressive disease phenotypes compared to the other groups. Altogether, these results suggest that, despite being regarded as a single disorder, AD can be characterized by different subtypes of brain atrophy, which might determine the progression of the disease.

There is increasing evidence showing that the spread of neurofibrillary tangles in the brain do not always conform to the Braak's staging scheme (Murray et al., 2011), which has been widely used to assess brain pathology at autopsy in AD. In particular, in a previous study (Murray et al., 2011), it was shown that 25% of AD cases present with an atypical pattern of tangles in the brain. This suggests that atypical presentations are frequent and should be considered in prospective longitudinal

studies. In line with these findings, our results showed that a similar proportion of patients (27.8%, n=83) did not present a diffuse pattern of brain atrophy, which is usually found at late stages of AD. Among these patients, 4% (n=12) were characterized by volume loss in medial temporal and limbic regions, including the hippocampus, anterior temporal poles, insula, orbitofrontal and anterior cingulate gyri. In addition, 5.7% (n=17) were characterized by atrophy in the parietal cortex, superior frontal and posterior temporal regions, with relative sparing of the medial temporal cortex and hippocampus. These two subtypes fit well with the description of atypical AD presentations shown in previous studies (Whitwell et al., 2012; Noh et al., 2012; Hwang et al., 2016), suggesting that structural MRI is a good biomarker to assess AD heterogeneity. The frequency of hippocampal-sparing patients we found in our study was similar to the one reported in previous studies (Hwang et al., 2016; Byun et al., 2015), while the frequency of the limbic predominant was somewhat different from previous studies probably to methodological approaches (Hwang et al., 2016; Byun et al., 2015; Ferreira et al., 2017). Moreover, we also found that 18.1% (n=54) of patients had minimal brain atrophy showing volume loss only in the left entorhinal region. This finding is in line with three recent studies showing that some AD patients do not show significant brain atrophy compared to controls (Byun et al., 2015; Dong et al., 2017; Ferreira et al., 2017), indicating that there might be a greater variability in brain atrophy patterns in AD than previously thought. Two of the recent studies that used the ADNI cohort (Byun et al., 2015; Ferreira et al., 2017), found that the levels of CSF $A\beta_{1-42}$ and total tau were abnormal in the minimal atrophy subtype and in between the typical and the limbic-predominamt subtypes. These findings indicate that the minimal atrophy subtype is more likely to have Alzheimer's disease etiology rather than depression or other psychiatric illnesses. The hippocampal-

sparing presented higher scores on all the ADAS-Cog subtests (Table 2) compared to the minimal atrophy patients, indicating that they were more impaired in all cognitive domains. The fact that these differences did not reach statistical significance (apart from the constructional praxis) could be due to the unbalanced sample sizes (54 MA versus 17 HS subjects).

Finally, in addition to the previous subtypes, in this study we also found two clusters of patients with diffuse brain atrophy. Although the two clusters had similar spatial patterns of atrophy, the diffuse 2 subtype presented greater frontal, parietal and temporal atrophy compared to the diffuse 1 subtype. Moreover, the diffuse 2 subtype performed significantly worse than the diffuse 1 both in general cognitive functions (CDR, MMSE) and memory (ADAS word recognition) as well as in orientation (ADAS orientation), praxis (ADAS ideational praxis) and language (ADAS following commands) components of the ADAS cognitive battery (Table 2).

In this study, the comparison of demographic variables between groups showed that the limbic-predominant and diffuse subtypes were significantly older than the other subtypes. Hence, their patterns of brain atrophy might be characteristic of AD patients with a more advanced age. It is well known that age has an impact on medial temporal lobe structures and that elderly individuals present greater atrophy in hippocampal and temporal regions (Launer et al., 1994; Raz et al., 2004; Pereira et al., 2014). These findings are in line with the temporal atrophy observed in the limbic-predominant AD patients in our study. These associations between increasing age and widespread volume losses in heteromodal association areas (Raji et al., 2009) could partially account for the severe brain atrophy we found in patients with diffuse atrophy subtypes, who were significantly older than patients from other groups.

Patients with the hippocampal-sparing subtype had a younger age at onset of AD. Several studies have shown that early onset AD is associated with gray matter atrophy in regions outside the medial temporal lobe (Frisoni et al., 2007; Shiino et al., 2008; Canu et al., 2012). These regions include the posterior and frontoparietal cortices and could be responsible for greater impairment in visuospatial or executive functions (van der Flier., 2011). There is also evidence showing that early onset AD might have a more aggressive disease course, showing a faster cognitive decline compared to the late onset cases (Van der Vlies, 2007). Our findings are in line with this notion since the hippocampal-sparing group showed a significant decline in visuospatial and memory functions after one year. The decline in visuospatial functions agrees with the widespread parietal atrophy these patients presented and the strong association that exists between the parietal cortex and visuospatial and visuoconstructional abilities (Cabeza and Nyberg, 2000). Previous studies using functional MRI (fMRI) have shown that memory not only relies on temporal brain areas but also the prefrontal cortex, which is important for episodic memory encoding and semantic retrieval (Cabeza and Nyberg, 2000). This region was atrophied in the hippocampal-sparing cases of our study, suggesting that it might be responsible for the memory decline they showed after one year.

In this study we found that patients with diffuse atrophy presented lower scores on the MMSE compared to the other groups, indicating that widespread brain atrophy was associated with greater overall cognitive impairment. In addition, they also showed a significant decline in visuospatial, memory and executive functions after one year, suggesting that, similarly to the hippocampal-sparing subtype, they might be a more aggressive disease phenotype. Two recent studies (Varol et al., 2016; Dong et al., 2017) also identified two diffuse atrophy subtypes, similarly to our study.

In a previous study, the ratio of hippocampus to cortex volumes provided a good discrimination between AD subtypes, suggesting it could be a good marker to assess AD heterogeneity (Whitwell et al., 2012). In our study, this ratio distinguished well the limbic-predominant, hippocampal-sparing and diffuse subtypes but not patients with the minimal atrophy subtype. This was because patients with minimal atrophy had the highest cortical volume, but their hippocampal volumes were lower than the hippocampal-sparing subtype. This is a numerical issue that arises from the ratio's formula and is not connected with the progression of the subtypes or their similarity. Since the AD subtypes atrophy patterns were not linear, it was not possible to distinguish them well using the previous ratio. Multiple anatomical brain measures and their non-linear relationships are needed to separate well different AD subtypes.

Regarding the supervised analysis, the precuneus, middle occipital and orbital gyri were the most important anatomical regions that helped discriminate clusters of patients (figure 4). The precuneus has been widely studied in AD and atrophy in this area has been associated with an early disease onset (Karas et al., 2007). The subtypes with the lowest precuneus volume were the hippocampal-sparing and the diffuse 2. The extent of atrophy in the hippocampal-sparing subtype can be explained by its onset, which is the earliest among the subtypes. As for the diffuse 2 subtype, the reason why the volume in this area was low is probably related to the severity of this subtype in contrast to the rest of the subtypes with respect to cortex volume (Figure 3, A). The volumes of the middle occipital gyrus separated the minimal-atrophy subtype from the other two subtypes of this branch and the interpretation of this is straightforward, since the minimal atrophy group showed no atrophy in that region. The orbital gyrus was the cortical area that distinguished the hippocampal-sparing and diffuses 2 subtypes. The hippocampal-sparing does not show atrophy in limbic

regions, including the orbital gyrus, and therefore its discrimination from the diffuse 2 subtype is feasible using this region.

Our study has a few limitations. The cross-sectional design did not allow us to examine anatomical changes over time. CSF measurements were absent for more than 50% of the sample; therefore we were not able to investigate the link between Aβ pathology and AD subtypes. The ROI approach in our analysis might have missed some subtle effects in cortical thinning.

Our study also has several strengths. For instance, the application of the random forest method in the cluster analyses is an important advantage compared to previous studies since it takes into account both the high dimensionality and the non-linear relationships that exist between anatomical regions. The inclusion of both cortical and subcortical regions in our analysis is also a strength that allowed us to contrast better our results with previous neuropathological studies assessing AD subtypes, including the one called limbic-predominant, which is characterized by hippocampal atrophy.

In conclusion, our study is the first to assess the patterns of cortical and subcortical atrophy in AD using two different cohorts in the same clustering analysis. Our findings suggest there is substantial heterogeneity among AD patients and that this heterogeneity has an important impact on how patients progress over time, even after a short follow-up period of one year. Future studies examining brain atrophy in AD should be aware that several subtypes of neurodegeneration might be present in their cohorts and that this can influence clinical presentation and prognosis.

5. Acknowledgements

We would also like to thank the Swedish Foundation for Strategic Research (SSF), The Swedish Research Council (VR), the Strategic Research Programme in

Neuroscience at Karolinska Institutet (StratNeuro), Swedish Brain Power, the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, Hjärnfonden, Alzheimerfonden, the Åke Wiberg Foundation and Birgitta och Sten Westerberg for additional financial support.

This study was supported by InnoMed, (Innovative Medicines in Europe) an Integrated Project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5, Life Sciences, Genomics and Biotechnology for Health.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and

Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

6. Disclosures

The authors have no conflicts of interest to declare.

7. References

Alzheimer's Association., 2016. 2016 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 12(4), 459-509.

Amit, Y., & Geman, D., 1997. Shape quantization and recognition with randomized trees. Neural computation, 9(7), 1545-1588.

Aoyan Dong et. al., 2016. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*; 140 (3): 735-747. doi: 10.1093/brain/aww319.

Banerjee, A., & Dave, R. N., 2004, July. Validating clusters using the Hopkins statistic. In Fuzzy systems, 2004. Proceedings. 2004 IEEE international conference on (Vol. 1, pp. 149-153). IEEE.

Blei, D. M., & Lafferty, J. D., 2007. A correlated topic model of science. The Annals of Applied Statistics, 17-35.

Braak, H., & Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. Acta neuropathologica, 82(4), 239-259 .

Breiman, L. Bagging predictors., 1996. Machine learning, 24(2), 123-140.

Breiman, L. Random forests., 2001. Machine learning, 45(1), 5-32.

Buckner, R. et al., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage, 23(2), 724-738.

Byun, M. S. et al., 2015. Heterogeneity of regional brain atrophy patterns associated with distinct progression rates in Alzheimer's disease. PloS one, 10(11), e0142756.

Cabeza, R., & Nyberg, L., 2006. Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of cognitive neuroscience, 12(1), 1-47.

Canu, E. et al., 2012. Early and late onset Alzheimer's disease patients have distinct patterns of white matter damage. Neurobiology of aging, 33(6), 1023-1033.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage, 53(1), 1-15.

Falahati F , Ferreira D, Soininen H, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Lovestone S, Eriksdotter M, Wahlund L.O. 2016. The effect of age correction on multivariate classification in Alzheimer's disease, with a focus on the characteristics of incorrectly and correctly classified subjects. Brain Topogr . 29:296–307.

Ferreira, D., et al., 2017. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Scientific Reports, 7:46263 | DOI: 10.1038/srep46263.

Fischl, B. et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron, 33(3), 341-355.

Fischl, B., & Dale, A. M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences, 97(20), 11050-11055.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human brain mapping, 8(4), 272-284.

Frisoni, G. B. et al., 2007. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain, 130(3), 720-730.

Frisoni, G. B. et al., 2010. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology, 6(2), 67-77.

Gray, K. et al., (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage, 65, 167-175.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M., 2001. On clustering validation techniques. Journal of intelligent information systems, 17(2-3), 107-145.

Hwang, J. et al., 2016. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2, 58-67.

Jack, C. R. et al., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of magnetic resonance imaging, 27(4), 685-691.

Karas, G. et al., 2007. Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. Neuroradiology, 49(12), 967-976.

Launer, L. J. et al., 1995. Medial temporal lobe atrophy in an open population of very old persons Cognitive, brain atrophy, and sociomedical correlates. Neurology, 45(4), 747-752.

Lovestone, S. et al., 2009. AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. Annals of the New York Academy of Sciences, 1180(1), 36-46.

Lovestone, S., Francis, P., & Strandgaard, K., 2007. Biomarkers for disease modification trials-The innovative medicines initiative and AddNeuroMed. The journal of nutrition, health & aging, 11(4), 359.

Muehlboeck, J., Westman, E., & Simmons, A., 2014. The HiveDB image data management and analysis framework. Frontiers in neuroinformatics, 7, 49.

Murray, M. E. et al., 2011. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. The Lancet Neurology, 10(9), 785-796.

Müller, A. C., & Guido, S. 2016. Introduction to machine learning with Python: a guide for data scientists. p85.

Na, H. K., Kang, D. R., Kim, S., Seo, S. W., Heilman, K. M., Noh, Y., & Na, D. L. 2016. Malignant progression in parietal-dominant atrophy subtype of Alzheimer's disease occurs independent of onset age. Neurobiology of aging, 47, 149-156.

Noh, Y. et al., 2014. Anatomical heterogeneity of Alzheimer disease based on cortical thickness on MRIs. Neurology, 83(21), 1936-1944.

O'Brien, L. M., 2006. Adjustment for whole brain and cranial size in volumetric brain studies: a review of common adjustment factors and statistical methods. Harvard review of psychiatry, 14(3), 141-151.

Park, J. Y. et al., 2017. Robust Identification of Alzheimer's Disease subtypes based on cortical atrophy patterns. Scientific Reports, 7: 43270 | DOI: 10.1038/srep43270.

Parsons, L., Haque, E., & Liu, H. 2004. Subspace clustering for high dimensional data: a review. *Acm Sigkdd Explorations Newsletter*, *6*(1), 90-105.

Passos, A., Wallach, H. M., & McCallum, A., 2011. Correlations and Anticorrelations in LDA Inference.

Pereira, J. B. et al., 2014. Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. Journal of internal medicine, 275(3), 317-330.

Pereira, J. B. et al., 2016. Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer's disease. Cerebral Cortex, 26(8), 3476-3493.

Raji, C. A. et al., 2009. Age, Alzheimer disease, and brain structure. Neurology, 73(22), 1899-1905.

Raz, N. et al., 2004. Differential aging of the medial temporal lobe a study of a five-year change. Neurology, 62(3), 433-438.

Ségonne, F., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage, 22(3), 1060-1075.

Ségonne, F., Pacheco, J., & Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE transactions on medical imaging, 26(4), 518-529.

Shaw, L. M. et al., 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Annals of neurology, 65(4), 403-413.

Shi, T., & Horvath, S. 2006. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics, 15*(1), 118-138.

Shiino, A. et al., 2008. Different atrophic patterns in early-and late-onset Alzheimer's disease and evaluation of clinical utility of a method of regional z-score analysis using voxel-based morphometry. Dementia and geriatric cognitive disorders, 26(2), 175-186.

Simmons, A. et al., 2011. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: experience from the first 24 months. International journal of geriatric psychiatry, 26(1), 75-82.

Simmons, A., 2009. MRI measures of Alzheimer's disease and the AddNeuroMed study. Annals of the New York Academy of Sciences, 1180(1), 47-55.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE transactions on medical imaging, 17(1), 87-97.

Spulber, G. et al., 2013. An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment. J Intern Med . 273:396–409.

van der Flier, W. M. et al., 2011. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ?4 allele. The Lancet Neurology, 10(3), 280-288.

van der Vlies, A. E., 2007. Cognitive impairment in Alzheimer's disease is modified by APOE genotype. Dementia and geriatric cognitive disorders, 24(2), 98-103.

Westman, E. et al., 2011. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. Neuroimage, 58(3), 818-828.

Whitwell, J. L. et al., 2012. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. The Lancet Neurology, 11(10), 868-877.

Voevodskaya, O. et al., 2014. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. Frontiers in aging neuroscience, 6, 264.

Zhang, X. et al., 2016. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proceedings of the National Academy of Sciences, 113(42), E6535-E6544.

Table 1. Differences between AD subtypes in subcortical volumes.

	Minimal atrophy	Limbic predominant	Hippocampal sparing	Diffuse 1	Diffuse 2	P-value[k]
Hippocampus [a b d e g h j]	6024 (969)	4566 (753)	5902 (846)	5252 (843)	5116 (772)	<0.001
Amygdala [a b d e h]	2299 (390)	1671 (386)	2134 (371)	1999 (372)	1962 (368)	<0.001
Thalamus [b d e g h j]	12380 (988)	11452 (1316)	12610 (982)	11980 (920)	11780 (885)	0.002
Caudate	6967 (942)	6430 (984)	7054 (1220)	6862 (1135)	6854 (1195)	0.520
Putamen [b e h]	9167 (1256)	7800 (1075)	8765 (1735)	8436 (1102)	8013 (1103)	<0.001
Accumbens[b d e h j]	749 (160)	547 (154)	703 (168)	648 (134)	595 (124)	<0.001
Pallidum[a b d h j]	2778 (284)	2460 (258)	2911 (350)	2708 (322)	2611 (292)	<0.001

The data are presented as mean (sd). All results and corresponding p values were corrected by multiple comparisons ($p < 0.05$):

[a] Significant differences between LP and D1 subtypes ($p < 0.05$)
[b] Significant differences between LP and MA subtypes ($p < 0.05$)
[c] Significant differences between LP and D2 subtypes ($p < 0.05$)
[d] Significant differences between LP and HS subtypes ($p < 0.05$)
[e] Significant differences between D1 and MA subtypes ($p < 0.05$)
[f] Significant differences between D1 and D2 subtypes ($p < 0.05$)
[g] Significant differences between D1 and HS subtypes ($p < 0.05$)
[h] Significant differences between MA and D2 subtypes ($p < 0.05$)
[I] Significant differences between MA and HS subtypes ($p < 0.05$)
[j] Significant differences between D2 and HS subtypes ($p < 0.05$)
[k] Significant differences across all AD groups ($p < 0.05$)

Table 2. Demographic and clinical characteristics of controls and different AD subtypes.

	Controls	Minimal atrophy	Limbic predominant	Hippocampal sparing	Diffuse 1	Diffuse 2	P value
N (%)	328	54 (18.1)	12 (4.0)	17 (5.7)	167 (55.8)	49 (16.4)	-
AddNeuroMed/ADNI N (%)		18/36 (15.8/19.5)	4/8 (3.5/4.3)	4/13 (3.5/7)	58/109 (50.8/58.9)	30/19 (26.3/10.3)	-
Women(%)	163 (49.7)	29 (54.0)	9 (75.0)	8 (47.0)	91 (54.5)	27 (55.1)	0.658
Age [b e h d g j]	74.9 (71.5-78)	72.0 (68.8-76)	80.0 (76.5-84)	70.9 (64-77)	77.0 (71-80.3)	79.0 (75-82)	<0.001
Disease duration [a b c]	-	3.00 (2-4.8)	4.50 (4.4-7.0)	4.00 (3-5)	3.00 (2-5)	2.60 (1.9-5)	0.073
Age at the Onset [e h g j]	-	69.5 (65-73)	75.0 (67.5-79)	67.0 (60-74)	73.0 (68-78)	75.0 (71-79)	<0.001
Education [j]	15.50 (12-18)	13.0 (8-16)	13.0 (9-15.5)	16.0 (10-18)	13.0 (10-16)	11.0 (5-15)	0.053
ApoE e4 (%)	88 (26.0)	36 (66.0)	6 (50.0)	10 (58.8)	101 (60.5)	28 (57.1)	0.845
MMSE [e f h]	29.0 (29-30)	24.0 (23-26)	22.5 (21.5-25)	23.0 (21-24)	23.0 (21-25)	21.0 (18-23)	<0.001
CDR general [b e f h j]	0.00 (0.0)	0.71 (0.3)	0.96 (0.1)	0.82 (0.39)	0.90 (0.41)	1.11 (0.5)	<0.001
Word recall task [e h]	3.70 (1.3)	5.50 (1.4)	6.20 (1.8)	6.00(1.1)	6.50 (1.5)	6.75 (1.3)	<0.001
Naming objects and fingers	-	0.40(0.7)	0.58 (0.7)	0.47 (0.8)	0.70(0.9)	0.73 (0.9)	0.140
Following commands [f h]	-	0.57 (0.7)	0.70(0.5)	0.90 (0.9)	0.80(1.0)	1.30(1.1)	0.004
Constructional praxis [h I]	-	0.70 (0.6)	0.75 (0.6)	1.50(1.0)	0.95 (0.8)	1.30(1.0)	0.005
Ideational praxis [f h]	-	0.26 (0.5)	0.75 (0.9)	0.94 (1.3)	0.50 (0.9)	1.10(1.0)	<0.001
Orientation total incorrect [b e f h j]	-	1.80 (1.8)	3.33 (2.0)	1.10 (1.5)	2.62 (2.0)	3.60 (2.0)	<0.001
Word recognition task mean incorrect [b e h]	-	5.00 (2.4)	7.75 (3.2)	5.70(2.9)	6.80 (3.1)	4.00(3.3)	0.001

Recall of test instruction [e f h]	-	0.24 (0.7)	0.50(1.0)	0.60(0.9)	0.66 (1.2)	1.10 (1.3)	<0.001
Spoken language ability [h]	-	0.30(0.6)	0.40 (0.7)	0.35 (1.0)	0.50 (0.8)	0.70 (0.9)	0.040
Word finding difficulty in spontaneous speech [c h]	-	0.60 (0.7)	0.25 (0.5)	0.90(1.1)	0.90 (1.1)	1.10 (1.0)	0.003

All values correspond to the median followed by range, except CDR and ADAS, in which they correspond to mean followed by standard deviation. ApoE e4, apolipoprotein allele e4; MMSE, mini-mental state examination; CDR, clinical dementia rating scale; ADAS, Alzheimer's disease assessment scale. P values were calculated using X^2 or Kruskal-Wallis tests across AD groups. The statistical tests are corrected for multiple comparisons (Holmes-Sidak) in 5% significance level. The AddNeuroMed/ADNI percentages are corrected for cohort population number.

[a] Significant differences between LP and D1 subtypes ($p < 0.05$)
[b] Significant differences between LP and MA subtypes ($p < 0.05$)
[c] Significant differences between LP and D2 subtypes ($p < 0.05$)
[d] Significant differences between LP and HS subtypes ($p < 0.05$)
[e] Significant differences between D1 and MA subtypes ($p < 0.05$)
[f] Significant differences between D1 and D2 subtypes ($p < 0.05$)
[g] Significant differences between D1 and HS subtypes ($p < 0.05$)
[h] Significant differences between MA and D2 subtypes ($p < 0.05$)
[I] Significant differences between MA and HS subtypes ($p < 0.05$)
[j] Significant differences between D2 and HS subtypes ($p < 0.05$)
[k] Significant differences across all AD groups ($p < 0.05$)

Table 3. Longitudinal comparisons in neuropsychological tests in the different AD subtypes.

	Minimal atrophy	Limbic predominant	Hippocampal sparing	Diffuse 1	Diffuse 2
Word recall task	5.35 (5,70)	6.37(6.59)	6.06 (6.92)*	6.40 (6.48)	6.38 (6.82)
Naming objects and fingers	0.44 (0,32)	0.44 (0.66)	0.50 (0.41)	0.54 (0.76)	0.64 (0.91)
Following commands	0.54 (0.98)	0.67 (1.11)	0.59 (1.04)	0.54 (0.96)*	0.74 (1.10)
Constructional praxis	0.71 (0.60)	0.78 (1.11)	0.86 (1.04)	0.92 (1.06)	1.07 (1.11)
Ideational praxis	0.22 (0.20)	0.78 (1.22)	0.82 (0.87)	0.41 (0.91)**	0.48 (1.05)***
Orientation total incorrect	1.51 (1.92)	3.67 (4.11)	1.82 (3.05)*	2.26 (3.38)***	2.64 (3.81)***
Word recognition task mean incorrect	4.88 (5.54)	8.22 (8.66)	5.73 (7.41)	6.79 (7.44)	6.24 (7.35)*
Recall of test instruction	0.27 (0.37)	0.33 (0.77)	0.36 (0.72)	0.41 (0.55)	0.76 (1.02)
Spoken language ability	0.22 (0.44)	0.44 (0.33)	0.45 (0.63)	0.46 (0.75)	0.53 (1.00)
Word finding difficulty in spontaneous speech	0.49 (0.69)	0.22 (0.88)	1.05 (1.41)	0.71 (1.00)	0.97 (1.37)
Comprehension of spoken language	0.29 (0.31)	0.33 (0.66)	0.50 (0.59)	0.45 (0.72)	0.47 (0.74)

The data are presented as mean on the baseline (mean on the 12 months follow up). The statistical tests are corrected for multiple comparisons (Holmes-Sidak) in 5% significance level.

*Significant (<0.05), **Significant (<0.0005),*** Significant (<0.00005)

For the follow up, a subsample of the original sample was available: MA=41 (76%), LP=9 (75%), HS=10 (59%), D1=121 (78%), D2=27 (55%).

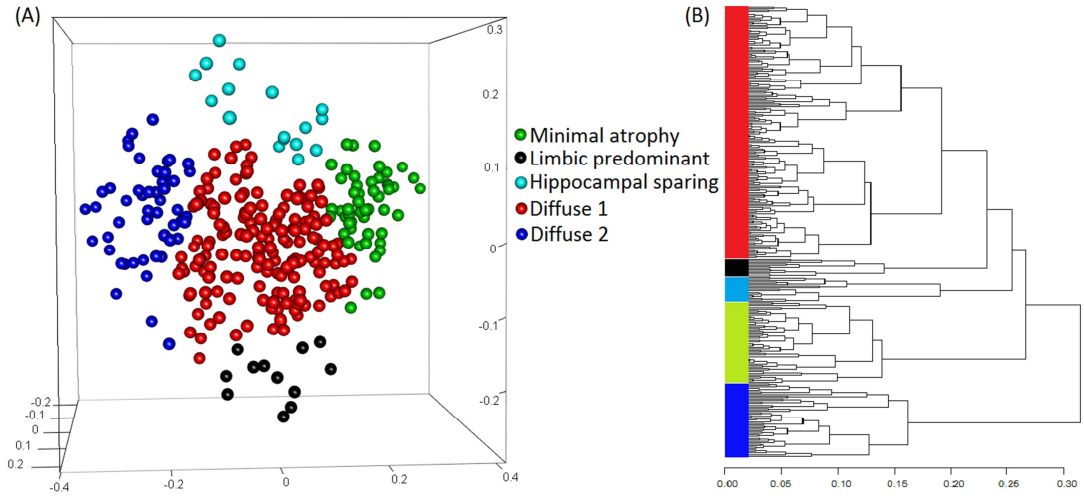


Fig. 1. Cluster analysis visual results in AD patients. (A) 3D Multidimensional scaling representation of the similarity matrix. Dots represent AD subjects, the distance between them symbolizes how similar the subjects are with respect to their cortical and subcortical patterns and the color denotes the clusters obtained by the hierarchical clustering. (B) Hierarchical clustering tree result. The horizontal axis represents the distance in random forest similarity between AD subjects, while the different colors of the vertical color bar correspond to different subtypes: Minimal atrophy (green), limbic predominant (black), hippocampal-sparing (light blue), diffuse 1 (red) and diffuse 2 (dark blue).

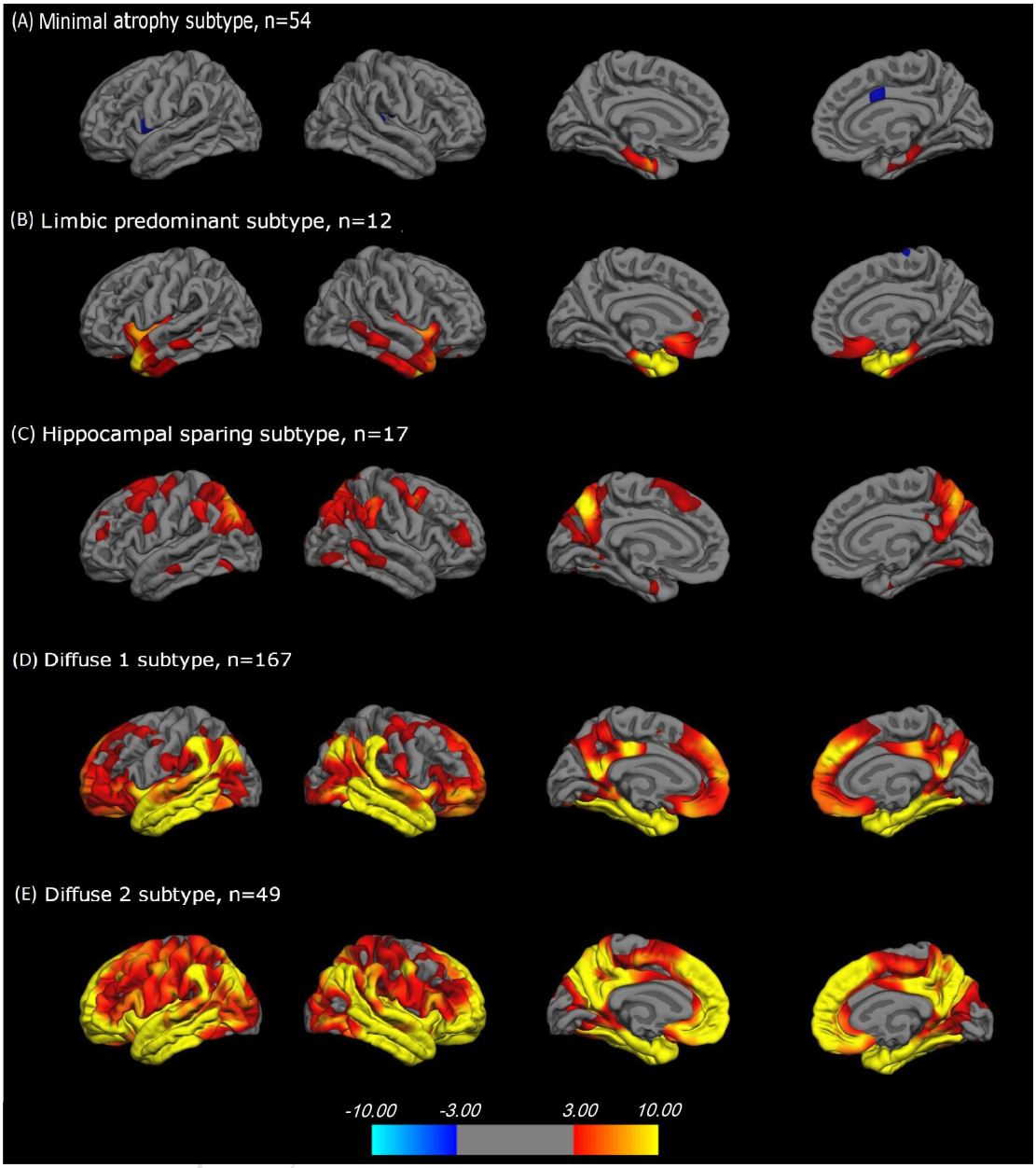


Fig. 2. Cortical volume patterns in different AD subtypes. Significant cortical volume loss in the minimal atrophy subtype (A), the limbic-predominant subtype (B), the hippocampal-sparing subtype (C), the diffuse 1 subtype (D) and the diffuse 2 subtype (E) compared to the CN group. Higher values (red/yellow color) indicate stronger differences between groups, threshold at p=0.001. The color bar denotes $-\log(p_{value})$.

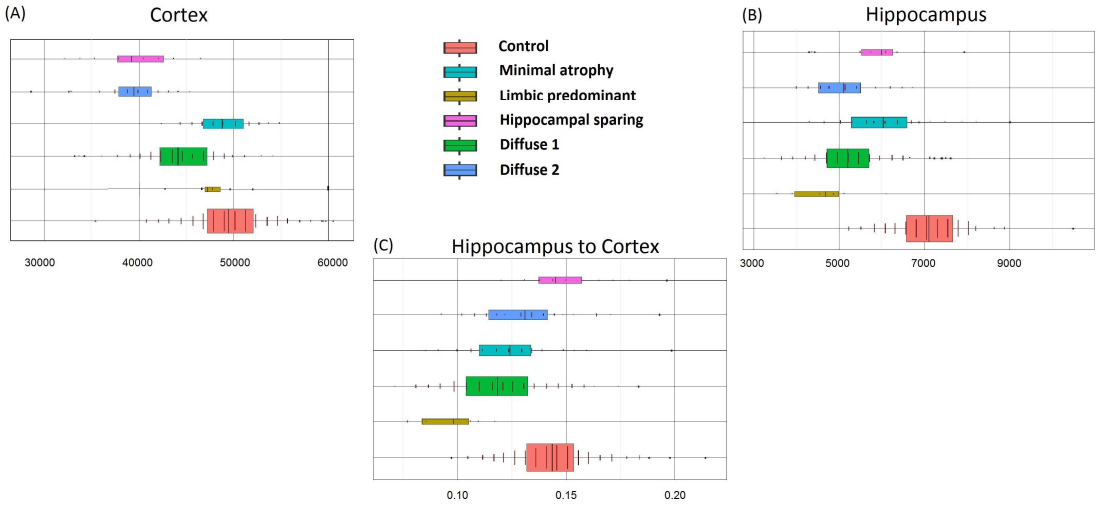


Fig. 3. Cortex, hippocampus and hippocampus to cortex volume ratio for the different AD subtypes and controls. For the five subtypes:
(A) shows the boxplots of the cortex volume and is measured in mm^3
(B) shows the boxplots of hippocampal volume and is measured in mm^3
(C) shows the ratio hippocampus to cortex.
The statistical differences of the cortex, hippocampus and hippocampus to cortex ratio between the subtypes are presented in the supplementary Table 3

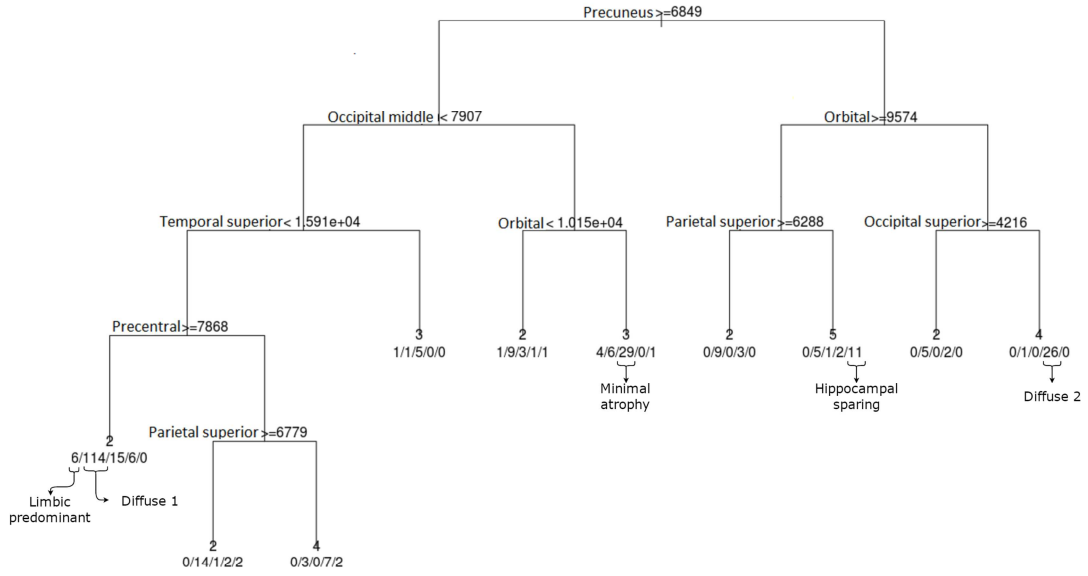


Fig. 4. Supervised classification tree with the most important variables and the subtypes of the random forest clustering. The 10 most important variables with decreasing accuracy were used in the training of this classification tree. The algorithm of the tree only keeps the variables that help it classify the subtypes optimally. Each of the 10 terminal nodes represent how many subjects of each subtype belong there under the tree rules of the branch and is encoded as: $(x_1, x_2, x_3, x_4, x_5)\rightarrow($ limbic predominant, diffuse 1, minimal atrophy, diffuse 2, hippocampal-sparing). The brackets indicate the terminal where most cases from each subtype were classified.

Subtypes of Alzheimer's disease can be found with high-dimensional cluster analysis.

Five distinct subtypes of Alzheimer's disease (AD) exhibit substantial heterogeneity.

Heterogeneity in AD has an impact on how patients function and progress over time.